(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number : **0 634 491 A1**

# (12) EUROPEAN PATENT APPLICATION

(21) Application number : **94870118.0**

(51) Int. Cl.[6] : **C12N 15/82**, C12N 15/54

(22) Date of filing : **11.07.94**

(30) Priority : **12.07.93 US 90523**

(43) Date of publication of application :
**18.01.95 Bulletin 95/03**

(84) Designated Contracting States :
**AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE**

(71) Applicant : **MONSANTO COMPANY**
**800 North Lindbergh Boulevard**
**St. Louis Missouri 63167 (US)**

(72) Inventor : **Barry, Gerard Francis**
**6350 Waterman Avenue**
**St. Louis, Missouri 63130 (US)**
Inventor : **Kishore, Ganesh Murthy**
**5354 Grantley Drive**
**Chesterfield, Missouri 63017 (US)**
Inventor : **Stark, David Martin**
**1218 Summerpoint Lane**
**Fenton, Missouri 63026 (US)**

(74) Representative : **Nash, Brian Walter et al**
**Monsanto Services International S.A.**
**Patent Department,**
**Avenue de Tervuren 270/272,**
**Letter Box No. 21**
**B-1150 Bruxelles (BE)**

## (54) Modified oil content in seeds.

(57) A method of decreasing the oil content of seeds by expression of ADPglucose pyrophosphorylase.

EP 0 634 491 A1

Recent advances in genetic engineering have provided the requisite tools to transform plants to contain foreign genes. It is now possible to produce plants which have unique characteristics of agronomic and crop processing importance. One such advantageous trait is enhanced starch and/or solids content and quality in various crop plants. WO 91/19806 reports the use of a gene which encodes ADPglucose pyrophosphorylase (ADPGPP), which catalyzes a key step in starch and glycogen biosynthesis. The preferred gene is from *E. coli* and the resulting enzyme is a poorly regulated, highly active variant.

Another desirable trait is the reduction of oil in certain food crops, such as peanuts. Decreasing the lipid content in the seeds of certain plants is desirable due to health concerns or for improved processing qualities. For example, a low calorie peanut butter, having a higher starch content and lower oil content would be beneficial. Also, soybeans having lower oil content would be better for producing certain products, such as tofu, soy sauce, soy meat extenders, and soy milk. In addition, lower oil content in certain seed-derived products is desirable, such as corn starch or wheat flour. It has surprisingly been found that such fat reduction is accomplished by expression of a gene encoding a deregulated ADPGPP, such as *glgC16* in the seeds.

## SUMMARY OF THE INVENTION

The present invention provides structural DNA constructs which encode an ADPglucose pyrophosphorylase (ADPGPP) enzyme and which are useful in producing seeds having a reduced oil content.

In accomplishing the foregoing, there is provided, in accordance with one aspect of the present invention, a method of producing genetically transformed plants which have elevated starch content, comprising the steps of:

(a) inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising
(i) a promoter which is selected from the group consisting of seed specific promoters,
(ii) a structural DNA sequence that causes the production of an RNA sequence which encodes a fusion polypeptide comprising an amino-terminal plastid transit peptide and an ADPglucose pyrophosphorylase enzyme,
(iii) a 3' non-translated DNA sequence which functions in plant cells to cause transcriptional termination and the addition of polyadenylated nucleotides to the 3' end of the RNA sequence;
(b) obtaining transformed plant cells; and
(c) regenerating from the transformed plant cells genetically transformed plants which have an elevated starch content.

In accordance with another aspect of the present invention, there is provided a recombinant, double-stranded DNA molecule comprising in sequence:

(a) a promoter which is selected from the group consisting of seed specific promoters;
(b) a structural DNA sequence that causes the production of an RNA sequence which encodes a fusion polypeptide comprising an amino-terminal plastid transit peptide and an ADPglucose pyrophosphorylase enzyme; and
(c) a 3' non-translated region which functions in plant cells to cause transcriptional termination and the addition of polyadenylated nucleotides to the 3' end of the RNA sequence, said promoter being heterologous with respect to the structural DNA.

There has also been provided, in accordance with another aspect of the present invention, transformed plant cells that contain DNA comprised of the above-mentioned elements (a), (b) and (c). In accordance with yet another aspect of the present invention, differentiated oilseed crop plants are provided that have decreased oil content in the seeds.

## DETAILED DESCRIPTION OF THE INVENTION

The expression of a plant gene which exists in double-stranded DNA form involves transcription of messenger RNA (mRNA) from one strand of the DNA by RNA polymerase enzyme, and the subsequent processing of the mRNA primary transcript inside the nucleus. This processing involves a 3' non-translated region which adds polyadenylate nucleotides to the 3' end of the RNA.

Transcription of DNA into mRNA is regulated by a region of DNA usually referred to as the promoter. The promoter region contains a sequence of bases that signals RNA polymerase to associate with the DNA, and to initiate the transcription of mRNA using one of the DNA strands as a template to make a corresponding complimentary strand of RNA.

Promoters which are known or are found to cause transcription of DNA in plant cells can be used in the present invention. Such promoters may be obtained from a variety of sources such as plants and plant viruses and include, but are not limited to, the enhanced CaMV35S promoter and promoters isolated from plant genes

such as ssRUBISCO genes. As described below, it is preferred that the particular promoter selected should be capable of causing sufficient expression to result in the production of an effective amount of ADPGPP enzyme to cause the desired decrease in oil content. Therefore, it is preferred to bring about expression of the ADPGPP gene in the seed tissues of the plant and throughout the seed development. The promoter chosen should have the desired tissue and developmental specificity. Those skilled in the art will recognize that the amount of ADPGPP needed to induce the desired decrease in oil content may vary with the type of plant and furthermore that too much ADPGPP activity may be deleterious to the plant. Therefore, promoter function should be optimized by selecting a promoter with the desired tissue expression capabilities and approximate promoter strength and selecting a transformant which produces the desired ADPGPP activity in the target tissues. This selection approach from the pool of transformants is routinely employed in expression of heterologous structural genes in plants since there is variation between transformants containing the same heterologous gene due to the site of gene insertion within the plant genome. (Commonly referred to as "position effect").

Promoters may be identified to be seed specific by screening a cDNA library of a plant seed for genes which are selectively or preferably expressed in seeds and then determine the promoter regions to obtain seed selective or seed enhanced promoters. It is believed that most of the enzymes involved in carbohydrate metabolism have seed-specific forms from which seed-specific promoters may be obtained. Examples of such enzymes are sucrose synthase, invertase, and ADPGPP (both subunits).

Several seed-specific promoters are well known. β-conglycinin (also known as the 7S protein) is one of the major storage proteins in soybean (*Glycine max*) (Tierney, 1987). Promoters from each of the genes for its three subunits may be used in the present invention. The β-subunit of β-conglycinin has been expressed, using its endogenous promoter, in the seeds of transgenic petunia and tobacco, showing that the promoter functions in a seed-specific manner in other plants (Bray, 1987). Example 1 below demonstrates the use of the α' subunit of this promoter with an ADPGPP in canola. The gene for the 11S storage protein of soybean is also known to be expressed in a seed-specific manner and its promoter may be used in the present invention.

Two seed-specific promoters from *Brassica napus* have been identified. They are the promoter for napin and the promoter for cruciferin (Murphy, 1989). The promoters for the genes encoding phaseolin (from beans) and oleosin (from rape, soybean, and others) are also useful in the present invention. (Zheng, 1993).

The zeins are a group of storage proteins found in maize endosperm. Genomic clones for zein genes have been isolated (Pedersen, 1982), and the promoters from these clones, including the 15 kD, 16 kD, 19 kD, 22 kD, 27 kD, and gamma genes, could also be used to express an ADPGPP gene in the seeds of maize and other plants. An endosperm-specific promoter of the 19 kD zein has been identified (Quattrocchio, 1990). Other promoters known to function in maize include the promoters for the following genes: *waxy, Brittle, Shrunken 2*, Branching enzymes I and II, starch synthases, debranching enzymes, oleosins, glutelins, and sucrose synthases.

Examples of promoters suitable for expression of an ADPGPP gene in wheat include those for the genes for the ADPGPP subunits, for the granule bound and other starch synthases, for the branching and debranching enzymes, for the embryogenesis-abundant proteins, for the gliadins, and for the glutenins. Examples of such promoters in rice include those for the genes for the ADPGPP subunits, for the granule bound and other starch synthases, for the branching enzymes, for the debranching enzymes, for sucrose synthases, and for the glutelins (Zheng, 1993). Examples of such promoters for barley include those for the genes for the ADPGPP subunits, for the granule bound and other starch synthases, for the branching enzymes, for the debranching enzymes, for sucrose synthases, for the hordeins, for the embryo globulins, and the aleurone specific proteins.

Promoters for genes encoding proteins other that for storage or carbohydrate metabolism may be found to be useful in the present invention. For example, the acyl carrier protein gene has a promoter known to function in a seed-specific manner. (Baerson, 1993).

The RNA produced by a DNA construct of the present invention also contains a 5' non-translated leader sequence. This sequence can be derived from the promoter selected to express the gene, and can be specifically modified so as to increase translation of the mRNA. The 5' non-translated regions can also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence. The present invention is not limited to constructs, as presented in the following examples, wherein the non-translated region is derived from the 5' non-translated sequence that accompanies the promoter sequence. Rather, the non-translated leader sequence can be derived from an unrelated promoter or coding sequence as discussed above.

The DNA constructs of the present invention also contain a structural coding sequence in double-stranded DNA form, which encodes a fusion polypeptide comprising an amino-terminal plastid transit peptide and an ADPGPP enzyme. The ADPGPP enzyme utilized in the present invention is preferably subject to reduced allosteric control in plants. Such an unregulated ADPGPP enzyme may be selected from known enzymes which exhibit unregulated enzymatic activity or can be produced by mutagenesis of native bacterial, or algal or plant ADPGPP enzymes as discussed in greater detail hereinafter. In some instances, the substantial differences

in the nature of regulators modulating the activity of the wild type ADPGPP enzyme permits the use of the wild type gene itself; in these instances, the concentration of the regulators within plant organelles will facilitate elicitation of significant ADPGPP enzyme activity.

Bacterial ADPglucose Pyrophosphorylases

The *E. coli* ADPGPP has been well characterized as a tightly regulated enzyme. The activator fructose 1,6-bisphosphate has been shown to activate the enzyme by increasing its $V_{max}$, and by increasing the affinity of the enzyme for its substrates (Preiss, 1966 and Gentner, 1967). In addition, fructose 1,6-bisphosphate (FBP) also modulates the sensitivity of the enzyme to the inhibitors adenosine-5'-monophosphate (AMP) and inorganic phosphate ($P_i$) (Gentner, 1968).

In 1981, the *E. coli* K12 ADPGPP gene (*glg*C), along with the genes for glycogen synthase and branching enzyme, were cloned, and the resulting plasmid was named pOP12 (Okita, 1981). The *glg*C gene, which was sequenced in 1983, contains 1293 bp (SEQ ID NO:1) and encodes 431 amino acids (SEQ ID NO:2) with a deduced molecular weight of 48,762 (Baecker, 1983).

The *glg*C16 gene was generated by chemically mutagenizing *E. coli* K12 strain PA 601 with N-methyl-N'-nitrosoguanidine (Cattaneo, 1969 and Creuzet-Sigal, 1972). When the kinetics of the *glg*C16 ADPGPP were compared to the parent, it was found that the *glg*C16 ADPGPP had a higher affinity for ADPglucose in the absence of the activator, fructose 1,6-bisphosphate (FBP), and the concentration of FBP needed for half-maximal activation of the enzyme was decreased in *glg*C16. The inhibition of the ADPGPP activity in *glg*C16 by 5'-AMP (AMP) was also reduced.

The DNA sequence of the *glg*C16 gene is now known (SEQ ID NO:3) (Kumar, 1989). When the *glg*C16 deduced amino acid sequence (SEQ ID NO:4) was compared to the nonisogenic *E. coli* K-12 3000, one amino acid change was noted: Gly 336 to Asp (Meyer et al., 1993).

A number of other ADPGPP mutants have been found in *E. coli*. The expression of any of these or other bacterial ADPGPP wild type or mutants could also be used to increase starch production in plants. *E. coli* K12 strain 6047 (*glg*C47) accumulates about the same amount of glycogen during stationary phase as does strain 618 (*glg*C16). Strain 6047, like 618, shows a higher apparent affinity for FBP, and more activity in the absence of FBP. However, the enzyme from strain 6047 is reportedly more sensitive to inhibition by AMP compared to the enzyme from strain 618 (Latil-Damotte, 1977).

The *glgC* gene from *Salmonella typhimurium* LT2 has also been cloned and sequenced (Leung and Preiss 1987a). The gene encodes 431 amino acids with a deduced molecular weight of 45,580. The *Salmonella typhimurium* LT2 *glgC* gene and the same gene from *E. coli* K-12 have 90% identity at the amino acid level and 80% identity at the DNA level. Like the *E. coli* ADPGPP, the *Salmonella typhimurium* LT2 ADPGPP is also activated by FBP and is inhibited by AMP (Leung and Preiss 1987b). This substantial conservation in amino acid sequences suggests that introduction of mutations which cause enhancement of ADPGPP activity in *E. coli* into *S. typhimurium* ADPGPP gene should have a similar effect on the ADPGPP enzyme of this organism.

A number of other bacterial ADPGPPs have been characterized by their response to activators and inhibitors (for review see: Preiss 1973). Like the *Escherichia coli* ADPGPP, the ADPGPPs from *Aerobacter aerogenes*, *Aerobacter cloacae*, *Citrobacter freundii*, and *Escherichia aurescens* are all activated by FBP and are inhibited by AMP. The ADPGPP from *Aeromonas formicans* is activated by fructose 6-phosphate or FBP, and is inhibited by ADP. The *Serratia marcescens* ADPGPP, however, was not activated by any metabolite tested. The photosynthetic *Rhodospirillum rubrum* has an ADPGPP that is activated by pyruvate, and none of the tested compounds, including $P_i$, AMP or ADP, inhibit the enzyme. Several algal ADPGPPs have been studied and found to have regulation similar to that found for plant ADPGPPs. Obviously, the ADPGPPs from many organisms could be used to increase starch biosynthesis and accumulation in plants.

Plant ADPglucose Pyrophosphorylases

At one time, UDPglucose was thought to be the primary substrate for starch biosynthesis in plants. However, ADPglucose was found to be a better substrate for starch biosynthesis than UDPglucose (Recondo, 1961). This same report states that ADPGPP activity was found in plant material.

A spinach leaf ADPGPP was partially purified and was shown to be activated by 3-phosphoglycerate (3-PGA) and inhibited by inorganic phosphate (Ghosh et al., 1966). The report by Ghosh et al. suggested that the biosynthesis of leaf starch was regulated by the level of ADPglucose. The activator, 3-PGA, is the primary product of $CO_2$ fixation in photosynthesis. During photosynthesis, the levels of 3-PGA would increase, causing activation of ADPGPP. At the same time, the levels of $P_i$ would decrease because of photophosphorylation, decreasing the inhibition of ADPGPP. These changes would cause an increase in ADPglucose production and

starch biosynthesis. During darkness, 3-PGA levels would decrease, and $P_i$ levels would increase, decreasing the activity of ADPGPP and, therefore, decreasing biosynthesis of ADPglucose and starch.

The ADPGPP from spinach leaves was later purified to homogeneity and shown to contain subunits of 51 and 54 kDa (Morell, 1987). Based on antibodies raised against the two subunits, the 51 kDa protein has homology with both the maize endosperm and potato tuber ADPGPPs, but not with the spinach leaf 54 kDa protein.

The sequence of a rice endosperm ADPGPP subunit cDNA clone has been reported (Anderson, 1989a). The clone encoded a protein of 483 amino acids. A comparison of the rice endosperm ADPGPP and the *E. coli* ADPGPP protein sequences shows about 30% identity. Also in 1989, an almost full-length cDNA clone for the wheat endosperm ADPGPP was sequenced (Olive, 1989). The wheat endosperm ADPGPP clone has about 24% identity with the *E. coli* ADPGPP protein sequence, while the wheat and the rice clones have 40% identity at the protein level.

The maize endosperm ADPGPP has been purified and shown to have catalytic and regulatory properties similar to those of other plant ADPGPPs (Plaxton, 1987). The native molecular weight of the maize endosperm enzyme is 230,000, and it is composed of four subunits of similar size.

The native molecular weight of the potato tuber ADPGPP is reported to be 200,000, with a subunit size of 50,000 (Sowokinos, 1982). Activity of the tuber ADPGPP is almost completely dependent on 3-PGA, and as with other plant ADPGPPs, is inhibited by $P_i$. The potato tuber and leaf ADPGPPs have been demonstrated to be similar in physical, catalytic, and allosteric properties (Anderson, 1989b).

Production of Altered ADPglucose Pyrophosphorylase Genes by Mutagenesis

Those skilled in the art will recognize that while not absolutely required, enhanced results are to be obtained by using ADPGPP genes which are subject to reduced allosteric regulation ("deregulated") and more preferably not subject to significant levels of allosteric regulation ("unregulated") while maintaining adequate catalytic activity. In cells which do not normally accumulate significant quantities of starch, expression of a "regulated" enzyme may be sufficient. In starch-accumulating cells and tissues, a "deregulated" or "unregulated" enzyme is the preferred system. The structural coding sequence for a bacterial or plant ADPGPP enzyme can be mutagenized in *E. coli* or another suitable host and screened for increased glycogen production as described for the *glg*C16 gene of *E. coli*. It should be realized that use of a gene encoding an ADPGPP enzyme which is only subject to modulators (activators/inhibitors) which are present in the selected plant at levels which do not significantly inhibit the catalytic activity will not require enzyme (gene) modification. These "unregulated" or "deregulated" ADPGPP genes can then be inserted into plants as described herein to obtain transgenic plants having increased starch content.

For example, any ADPGPP gene can be cloned into the *E. coli* B strain AC70R1-504 (Leung, 1986). This strain has a defective ADPGPP gene, and is derepressed five- to seven-fold for the other glycogen biosynthetic enzymes. The ADPGPP gene/ cDNA's can be put on a plasmid behind the *E. coli glg*C promoter or any other bacterial promoter. This construct can then be subjected to either site-directed or random mutagenesis. After mutagenesis, the cells would be plated on rich medium with 1% glucose. After the colonies have developed, the plates would be flooded with iodine solution (0.2 w/v% $I_2$, 0.4 w/v% KI in $H_2O$, Creuzet-Sigal, 1972). By comparison with an identical plate containing non-mutated *E. coli*, colonies that are producing more glycogen can be detected by their darker staining.

Since the mutagenesis procedure could have created promoter mutations, any putative ADPGPP mutant from the first round screening will have to have the ADPGPP gene recloned into non-mutated vector and the resulting plasmid will be screened in the same manner. The mutants that make it though both rounds of screening will then have their ADPGPP activities assayed with and without the activators and inhibitors. By comparing the mutated ADPGPP's responses to activators and inhibitors to the non-mutated enzymes, the new mutant can be characterized.

The report by Plaxton and Preiss in 1987 demonstrates that the maize endosperm ADPGPP has regulatory properties similar to those of the other plant ADPGPPs. They show that earlier reports claiming that the maize endosperm ADPGPP had enhanced activity in the absence of activator (3-PGA) and decreased sensitivity to the inhibitor ($P_i$), was due to proteolytic cleavage of the enzyme during the isolation procedure. By altering an ADPGPP gene to produce an enzyme analogous to the proteolytically cleaved maize endosperm ADPGPP, decreased allosteric regulation will be achieved. The recent report concerning the apparent novelty of the regulation of the barley endosperm ADPGPP and its apparent insensitivity to 3-PGA is not generally accepted since the report shows that the enzyme preparation was rapidly degraded and may suffer from the same problems identified for the corn endosperm preparation.

To assay a liquid culture of *E. coli* for ADPGPP activity, the cells are spun down in a centrifuge and resuspended in about 2 ml of extraction buffer (0.05 M glycylglycine pH 7.0, 5.0 mM DTE, 1.0 mM EDTA) per

gram of cell paste. The cells are lysed by passing twice through a French Press. The cell extracts are spun in a microcentrifuge for 5 minutes, and the supernatants are desalted by passing through a G-50 spin column.

The enzyme assay for the synthesis of ADPglucose is a modification of a published procedure (Haugen, 1976). Each 100 µl assay contains: 10 µmole Hepes pH 7.7, 50 µg BSA, 0.05µmole of [$^{14}$C]glucose-1-phosphate, 0.15 µmole ATP, 0.5 µmole $MgCl_2$, 0.1 µg of crystalline yeast inorganic pyrophosphatase, 1 mM ammonium molybdate, enzyme, activators or inhibitors as desired, and water. The assay is incubated at 37°C for 10 minutes, and is stopped by boiling for 60 seconds. The assay is spun down in a microcentrifuge, and 40 µl of the supernatant is injected onto a Synchrom Synchropak AX-100 anion exchange HPLC column. The sample is eluted with 65 mM KPi pH 5.5. Unreacted [$^{14}$C]glucose-1-phosphate elutes around 7-8 minutes, and [$^{14}$C]ADPglucose elutes at approximately 13 minutes. Enzyme activity is determined by the amount of radioactivity found in the ADPglucose peak.

The plant ADPGPP enzyme activity is tightly regulated, by both positive (3-phosphoglycerate; 3-PGA) and negative effectors (inorganic phosphate; $P_i$) (Ghosh and Preiss, 1966; Copeland and Preiss 1981; Sowokinos and Preiss 1982; Morell et al., 1987; Plaxton and Preiss, 1987; Preiss, 1988;) and the ratio of 3-PGA:$P_i$ plays a prominent role in regulating starch biosynthesis by modulating the ADPGPP activity (Kaiser and Bassham, 1979). The plant ADPGPP enzymes are heterotetramers of two large/"shrunken" and two small/"Brittle" subunits (Morell et al., 1987; Lin et al., 1988a, 1988b; Krishnan et al., 1986; Okita et al., 1990) and there is strong evidence to suggest that the heterotetramer is the most active form of ADPGPP. Support for this suggestion comes from the isolation of plant "starchless" mutants that are deficient in either of the subunits (Dickinson and Preiss, 1969; Lin et al., 1988a, 1988b) and from the characterization of an "ADPGPP" homotetramer of small subunits that was found to have only low enzyme activity (Lin et al., 1988b). In addition, proposed effector interaction residues have been identified for both subunits (Morell et al., 1988). Direct evidence for the active form of the enzyme and further support of the kinetic data reported for the purified potato enzyme comes from the expression of potato ADPGPP activity in *E. coli* and the comparison of the kinetic properties of this material and that from potato tubers (Iglesias et al., 1993).

Unregulated enzyme variants of the plant ADPGPP are identified and characterized in a manner similar to that which resulted in the isolation of the *E. coli glgC16* and related mutants. A number of plant ADPGPP cDNA's, or portions of such cDNA's, for both the large and small subunits, have been cloned from both monocots and dicots (Anderson et al., 1989a; Olive et al., 1989; Muller et al., 1990) The proteins encoded by the plant cDNA's, as well as those described from bacteria, show a high degree of conservation (Bhave et al., 1990). In particular, a highly conserved region, also containing some of the residues implicated in enzyme function and effector interactions, has been identified (Morell et al., 1988; Smith-White and Preiss, 1992). Clones of the potato tuber ADPGPP subunit genes have been isolated. These include a complete small subunit gene, assembled by addition of sequences from the first exon of the genomic clone with a nearly full-length cDNA clone of the same gene, and an almost complete gene for the large subunit. The nucleotide sequence (SEQ ID NO:7) and the amino acid sequence (SEQ ID NO:8) of the assembled small subunit gene are given below. The nucleotide sequence presented here differs from the gene originally isolated in the following ways: a *BglII+NcoI* site was introduced at the ATG codon to facilitate the cloning of the gene into *E. coli* and plant expression vectors by site directed mutagenesis utilizing the oligonucleotide primer sequence

GTTGATAACAAGATCTGTTAACCATGGCGGCTTCC (SEQ ID NO:11).

A *SacI* site was introduced at the stop codon utilizing the oligonucleotide primer sequence

CCAGTTAAAACGGAGCTCATCAGATGATGATTC (SEQ ID NO:12).

The *SacI* site serves as a 3' cloning site. An internal *BglII* site was removed utilizing the oligonucleotide primer sequence

GTGTGAGAACATAAATCTTGGATATGTTAC (SEQ ID NO:13).

This assembled gene was expressed in E. coli under the control of the *recA* promoter in a *PrecA-gene10L* expression cassette (Wong et al., 1988) to produce measurable levels of the protein. An initiating methionine codon is placed by site-directed mutagenesis utilizing the oligonucleotide primer sequence

GAATTCACAGGGCCATGGCTCTAGACCC (SEQ ID NO:14)

to express the mature gene.

The nucleotide sequence (SEQ ID NO:9) and the amino acid sequence (SEQ ID NO:10) of the almost complete large subunit gene are given below. An initiating methionine codon has been placed at the mature N-terminus by site-directed mutagenesis utilizing the oligonucleotide primer sequence

AAGATCAAACCTGCCATGGCTTACTCTGTGATCACTACTG (SEQ ID NO:15).

The purpose of the initiating methionine is to facilitate the expression of this large subunit gene in *E. coli*. A *Hind*III site is located 103 bp after the stop codon and serves as the 3' cloning site. The complete large ADPGPP gene is isolated by the 5' RACE procedure (Rapid Amplification of cDNA Ends; Frohman, 1990; Loh, 1989). The oligonucleotide primers for this procedure are as follows:

1) GGGAATTCAAGCTTGGATCCCGGGCCCCCCCCCCCCCC (SEQ ID NO:16);

2) GGGAATTCAAGCTTGGATCCCGGG (SEQ ID NO:17); and

3) CCTCTAGACAGTCGATCAGGAGCAGATGTACG (SEQ ID NO:18).

The first two are the equivalent to the ANpolyC and the AN primers of Loh et al. (1989), respectively, and the third is the reverse complement to a sequence in the large ADPGPP gene, located after the *Pst* I site in SEQ ID NO:9. The PCR 5' sequence products are cloned as *Eco*RI/*Hind*III/*Bam*HI-*Pst*I fragments and are easily assembled with the existing gene portion.

The weakly regulated enzyme mutants of ADPGPP are identified by initially scoring colonies from a mutagenized *E. coli* culture that show elevated glycogen synthesis, by iodine staining of 24-48 hour colonies on Luria-Agar plates containing glucose at 1%, and then by characterizing the responses of the ADPGPP enzymes from these isolates to the positive and negative effectors of this activity (Cattaneo et al., 1969; Preiss et al., 1971). A similar approach is applied to the isolation of such variants of the plant ADPGPP enzymes. Given an expression system for each of the subunit genes, mutagenesis of each gene is carried out separately, by any of a variety of known means, both chemical or physical (Miller, 1972) on cultures containing the gene or on purified DNA. Another approach is to use a PCR procedure (Ehrlich, 1989) on the complete gene in the presence of inhibiting $Mn^{++}$ ions, a condition that leads to a high rate of misincorporation of nucleotides. A PCR procedure may also be used with primers adjacent to just a specific region of the gene, and this mutagenized fragment then recloned into the non-mutagenized gene segments. A random synthetic oligo-nucleotide procedure may also be used to generate a highly mutagenized short region of the gene by mixing of nucleotides in the synthesis reaction to result in misincorporation at all positions in this region. This small region is flanked by restriction sites that are used to reinsert this region into the remainder of the gene. The resultant cultures or transformants are screened by the standard iodine method for those exhibiting glycogen levels higher than controls. Preferably this screening is carried out in an *E. coli* strain deficient only in ADPGPP activity and is phenotypically glycogen-minus and that is complemented to glycogen-plus by *glg*C. The *E. coli* strain should retain those other activities required for glycogen production. Both genes are expressed together in the same *E. coli* host by placing the genes on compatible plasmids with different selectable marker genes, and these plasmids also have similar copy numbers in the bacterial host to maximize heterotetramer formation. An example of such an expression system is the combination of pMON17335 and pMON17336 (Iglesias et al., 1993). The use of separate plasmids enables the screening of a mutagenized population of one gene alone, or in conjunction with the second gene following transformation into a competent host expressing the other gene, and the screening of two mutagenized populations following the combining of these in the same host. Following re-isolation of the plasmid DNA from colonies with increased iodine staining, the ADPGPP coding sequences are recloned into expression vectors, the phenotype verified, and the ADPGPP activity and its response to the effector molecules determined. Improved variants will display increased $V_{max}$, reduced inhibition by the negative effector ($P_i$), or reduced dependence upon activator (3-PGA) for maximal activity. The assay for such improved characteristics involves the determination of ADPGPP activity in the presence of $P_i$ at 0.045 mM ($I_{0.5}$ = 0.045 mM) or in the presence of 3-PGA at 0.075 mM ($A_{0.5}$ = 0.075 mM). The useful variants will display <40% inhibition at this concentration of $P_i$ or display >50% activity at this concentration of 3-PGA. Following the isolation of improved variants and the determination of the subunit or subunits responsible, the mutation(s) are

determined by nucleotide sequencing. The mutation is confirmed by recreating this change by site-directed mutagenesis and reassay of ADPGPP activity in the presence of activator and inhibitor. This mutation is then transferred to the equivalent complete ADPGPP cDNA gene, by recloning the region containing the change from the altered bacterial expression form to the plant form containing the amyloplast targeting sequence, or by site-directed mutagenesis of the complete native ADPGPP plant gene.

Chloroplast/Amyloplast Directed Expression of ADPGPP Activity

Starch biosynthesis is known to take place in plant chloroplasts and amyloplasts (herein collectively referred to as plastids. In the plants that have been studied, the ADPGPP is localized to these plastids. Many chloroplast-localized proteins are expressed from nuclear genes as precursors and are targeted to the chloroplast by a chloroplast transit peptide (CTP) that is removed during the import steps. Examples of such chloroplast proteins include the small subunit of Ribulose-1,5-bisphosphate carboxylase (ssRUBISCO, SSU), 5-enolpyruvateshikimate-3-phosphate synthase (EPSPS), Ferredoxin, Ferredoxin oxidoreductase, the Light-harvesting-complex protein I and protein II, and Thioredoxin F. It has been demonstrated *in vivo* and *in vitro* that non-chloroplast proteins may be targeted to the chloroplast by use of protein fusions with a CTP and that a CTP sequence is sufficient to target a protein to the chloroplast. Likewise, amyloplast-localized proteins are expressed from nuclear genes as precursors and are targeted to the amyloplast by an amyloplast transit peptide (ATP).

In the exemplary embodiments, a specialized CTP, derived from the ssRUBISCO 1A gene from *Arabidopsis thaliana* (SSU 1A) (Timko, 1988) was used. This CTP (CTP1) was constructed by a combination of site-directed mutageneses. The CTP1 nucleotide sequence (SEQ ID NO:5) and the corresponding amino acid sequence (SEQ ID NO:6) are given below. CTP1 is made up of the SSU 1A CTP (amino acid 1-55), the first 23 amino acids of the mature SSU 1A protein (56-78), a serine residue (amino acid 79), a new segment that repeats amino acids 50 to 56 from the CTP and the first two from the mature protein (amino acids 80-87), and an alanine and methionine residue (amino acid 88 and 89). An *Nco*I restriction site is located at the 3' end (spans the Met codon) to facilitate the construction of precise fusions to the 5' of an ADPGPP gene. At a later stage, a *Bgl*II site was introduced upstream of the N-terminus of the SSU 1A sequences to facilitate the introduction of the fusions into plant transformation vectors. A fusion was assembled between the structural DNA encoding the CTP1 CTP and the *glgC16* gene from *E. coli* to produce a complete structural DNA sequence encoding the plastid transit peptide/ADPGPP fusion polypeptide.

Those skilled in the art will recognize that if either a single plant ADPGPP cDNA encoding shrunken and/or brittle subunits or both plant ADPGPP cDNA's encoding shrunken and brittle subunits is utilized in the practice of the present invention, the endogenous CTP or ATP could most easily and preferably be used. Hence, for purposes of the present invention the term "plastid transit peptides" should be interpreted to include both chloroplast transit peptides and amyloplast transit peptides. Those skilled in the art will also recognize that various other chimeric constructs can be made which utilize the functionality of a particular plastid transit peptide to import the contiguous ADPGPP enzyme into the plant cell chloroplast/amyloplast depending on the promoter tissue specificity.

Polyadenylation Signal

The 3' non-translated region of the chimeric plant gene contains a polyadenylation signal which functions in plants to cause the addition of polyadenylate nucleotides to the 3' end of the RNA. Examples of suitable 3' regions are (1) the 3' transcribed, non-translated regions containing the polyadenylated signal of *Agrobacterium* the tumor-inducing (Ti) plasmid genes, such as the nopaline synthase (NOS) gene, and (2) plant genes like the soybean storage protein genes and the small subunit of the ribulose-1,5-bisphosphate carboxylase (ssRUBISCO) gene. An example of a preferred 3' region is that from the NOS gene, described in greater detail in the examples below.

Plant Transformation/Regeneration

Plants which can be made to have decreased oil content by practice of the present invention include, but are not limited to, corn, wheat, rice, pea, peanut, canola/oilseed rape, cotton, barley, sorghum, soybean, sunflower, almond, cashew, pecan, and walnut.

A double-stranded DNA molecule of the present invention containing the functional plant ADPGPP gene can be inserted into the genome of a plant by any suitable method. Suitable plant transformation vectors include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed, e.g., by Herrera-

Estrella (1983), Bevan (1983), Klee (1985) and EPO publication 120,516 (Schilperoort et al.). In addition to plant transformation vectors derived from the Ti or root-inducing (Ri) plasmids of *Agrobacterium*, alternative methods can be used to insert the DNA constructs of this invention into plant cells. Such methods may involve, for example, the use of liposomes, electroporation, chemicals that increase free DNA uptake, free DNA delivery via microprojectile bombardment, and transformation using viruses or pollen.

Examples of vectors designed for the expression of *glg*C16 and other ADPGPP genes in monocots and dicots are reported by Kishore in WO 91/19806. These are used to transform the desired plant cells by the appropriate method.

When adequate numbers of cells (or protoplasts) containing the ADPGPP gene or cDNA are obtained, the cells (or protoplasts) are regenerated into whole plants. Choice of methodology for the regeneration step is not critical, with suitable protocols being available for hosts from Leguminosae (alfalfa, soybean, clover, etc.), Cruciferae (cabbage, radish, canola/rapeseed, etc.), Gramineae (wheat, barley, rice, corn, etc.), various floral crops, such as sunflower, and nut-bearing trees, such as almonds, cashews, walnuts, and pecans. See, e.g., Ammirato, 1984; Shimamoto, 1989; Fromm, 1990; Vasil, 1990; Hayashimoto, 1989; and Datta, 1990.

The following examples are provided to better elucidate the practice of the present invention and should not be interpreted in any way to limit the scope of the present invention. Those skilled in the art will recognize that various modifications, truncations, etc. can be made to the methods and genes described herein while not departing from the spirit and scope of the present invention.

## Example 1

To express the *E. coli glg*C16 gene in plant cells, and to target the enzyme to the plastids, the gene needed to be fused to a DNA encoding the plastid-targeting transit peptide (hereinafter referred to as the CTP/ADPGPP gene), and to the proper plant regulatory regions. Detailed examples of how to accomplish this may be found in WO 91/19806.

The CTP-*glg*C16 gene fusion was placed behind the soybean β-conglycinin 7S storage promoter described above. This cassette was cloned into pMON17227, a Ti plasmid vector disclosed and described by Barry et al. in WO 92/04449 (1991), to form the vector pMON17315. This vector was used to transform canola by *Agrobacterium* transformation followed by glyphosate selection. Regenerated plants were analyzed and the presence of the enzyme in most transformants was confirmed by Western blot analysis. Seeds from four transformed lines have been obtained and analyzed for oil, starch, and protein content and moisture. The starch content was found to have increased to 8.2-18.2 percent (based on fresh weight) as compared to 0.9-1.6 percent in control lines (transformed with pMON17227 only). The oil content was found to have been decreased from 26.7-31.6 percent in the controls to 13.0-15.5 percent in the transformed lines. Protein content and moisture were not significantly changed. In some lines seed weight was increased which may indicate that total yield may also be increased.

## REFERENCES


Ammirato, P.V., et al. Handbook of Plant Cell Culture - Crop Species. Macmillan Publ. Co. (1984).

Anderson, et al. (1989a) J. Biol. Chem. 264 (21):12238-12242.

Anderson, et al. (1989b) First International Symposium on the Molecular Biology of the Potato, Bar Harbor, Maine.

Baecker, et al. (1983) J. Biol. Chem. 258 (8):5084-5087.

Baerson, et al. (1993) Plant Mol. Biol. 22:255-267.

Bevan, M. (1984) Nucleic Acids Res. 12 (22): 8711-8721.

Bhave, et al. (1990) Plant Cell 2: 581-588.

Bray, et al. (1987) Planta 172: 364-370.

Cattaneo, et al. (1969) Biochem. Biophys. Res. Commun. 34(5):694-701.

Chen et al. (1986) PNAS - US 83:8560-64.

Copeland, L. and J. Preiss (1981) Plant Physiol. 68: 996-1001.

Creuzet-Sigal, et al. (1972) "Genetic Analysis and Biochemical Characterization of Mutants Impairing Glycogen Metabolism in Escherichia coli K12." In Biochemistry of the Glycosidic Linkage; An Integrated View. Ed. by R. Piras and H. G. Pontis. 647-680. New York: Academic Press Inc.

Datta, et al. (1990) Biotechnology 8:736-740.

Dickinson, D. B. and J. Preiss (1969) Plant Physiol. 44:1058-1062.

Doyle, et al. (1986) J. Biol. Chem. 261(20):9228-38.

Ehrlich, H. A. (1989) Ed. PCR Technology - Principles and Applications for DNA Amplification. Stockton Press,

New York.
Frohman, et al. (1988) PNAS USA 85: 8998-9002.
Fromm et al. (1990) Bio/Technology 8: 833-839.
Gentner, et al. (1967) Bioch. Biophys. Res. Commun. 27(3):417-423.
Gentner, et al. (1968) J. Biol. Chem. 243(22):5882-5891.
Ghosh, et al. (1966) J. Biol. Chem. 241(19):4491-4504.
Haugen, et al. (1976) J. Biol. Chem. 251(24): 7880-7885.
Hayashimoto, et al. (1990) Plant Physiol. 93:857-863.
Herrera-Estrella, L., et al. (1983) Nature 303:209.
Iglesias, et al., (1993) J. Biol. Chem. 268:1081-1086.
Kaiser, et al. (1979) Plant Physiol. 63: 109-113.
Klee, H.J., et al. (1985) Bio/Technology 3:637-42.
Kumar, et al. (1989) J. Biol. Chem. 264(18): 10464-10471.
Latil-Damotte, et al. (1977) Mol. Gen. Genet. 150:325-329.
Leung, et al. (1986) J. Bact. 167(1):82-88.
Leung, et al. (1987a) J. Bact. 169 (9):4355-4360.
Leung, et al. (1987b) J. Bact. 169 (9):4349-4354.
Lin, et al. (1988a) Plant Physiol. 86:1131-1135.
Lin, et al. (1988b) Plant Physiol. 88:1175-1181.
Loh, et al. (1989) Science 243:217-220.
Meyer et al. (1993) Arch. Biochemistry & Biophys. 302:64-71.
Miller, J. H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Morell, et al. (1988) J. Biol. Chem. 263:633-637.
Morell, et al. (1987) Plant Physiol. 85:182-187.
Muller, et al. (1990) Mol. Gen. Genet. 224:136-146.
Murphy, et al. (1989) Plant Physiol. 135:63-69.
Okita, et al. (1981) J. Biol. Chem. 256(13):6944-6952.
Okita, et al. (1990) Plant Physiol. 93: 785-790.
Olive, et al. (1989) Plant Mol. Biol. 12:525-538.
Pedersen, et al. (1982) Cell 29:1015-1026.
Plaxton, W. C., and J. Preiss. (1987) Plant Physiol. 83:105-112.
Preiss, Jack. (1973) Adenosine Diphosphoryl Glucose Pyrophosphorylase. In Group Transfer. Edited by P. D. Boyer. 73-119. New York: Academic Press.
Preiss, Jack. (1988) Biosynthesis of Starch and Its Regulation. In The Biochemistry of Plants. Edited by J. Preiss. 184-249. Orlando, FL: Academic Press.
Preiss, et al. (1966) Biochem. 5(6):1833-1845.
Preiss, et al. (1971) Biochem. Biophys. Res. Commun. 42: 180-186.
Quattrocchio, et al. (1990) Plant Mol. Biol. 15:81-93.
Recondo, et al. (1961) Bioch. Biophys. Res. Commun. 6(2):85-88.
Shimamoto, K. et al. (1989) Nature 338:274-276.
Smith-White, B. and J. Preiss (1992) J. Mol. Evol. 34:449-464.
Sowokinos, et al. (1982) Plant Physiol. 69:1459-1466.
Tierney, et al. (1987) Planta 172:356-363.
Timko, et al. (1988) "Genetic Engineering of Nuclear Encoded Components of the Photosynthetic Apparatus in Arabidopsis." In Impact of Chemistry on Biotechnology-A Multidisciplinary Discussion. Ed. by Phillips et al. 279-295. Washington DC: ACS Books.
Vasil, et al. (1990) Bio/Technology 8:429-434.
Wong, et al. (1988) Gene 68: 193-203.
Zheng, et al. (1993) The Plant Journal 4(2):357-366.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Monsanto Company
(B) STREET: 800 North Lindbergh Boulevard
(C) CITY: St. Louis
(D) STATE: Missouri
(E) COUNTRY: United States of America
(F) POSTAL CODE (ZIP): 63167
(G) TELEPHONE: (314)694-3131
(H) TELEFAX: (314)694-5435

(ii) TITLE OF INVENTION: Modified Oil Content in Seeds

(iii) NUMBER OF SEQUENCES: 18

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/090523
(B) FILING DATE: 12-JUL-1993

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/709663
(B) FILING DATE: 07-JUN-1991

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/539763
(B) FILING DATE: 18-JUN-1990

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1296 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1296

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1293

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATG GTT AGT TTA GAG AAG AAC GAT CAC TTA ATG TTG GCG CGC CAG CTG     48
Met Val Ser Leu Glu Lys Asn Asp His Leu Met Leu Ala Arg Gln Leu
 1               5                  10                  15

CCA TTG AAA TCT GTT GCC CTG ATA CTG GCG GGA GGA CGT GGT ACC CGC     96
Pro Leu Lys Ser Val Ala Leu Ile Leu Ala Gly Gly Arg Gly Thr Arg
            20                  25                  30

CTG AAG GAT TTA ACC AAT AAG CGA GCA AAA CCG GCC GTA CAC TTC GGC    144
Leu Lys Asp Leu Thr Asn Lys Arg Ala Lys Pro Ala Val His Phe Gly
        35                  40                  45

GGT AAG TTC CGC ATT ATC GAC TTT GCG CTG TCT AAC TGC ATC AAC TCC    192
Gly Lys Phe Arg Ile Ile Asp Phe Ala Leu Ser Asn Cys Ile Asn Ser
    50                  55                  60

GGG ATC CGT CGT ATG GGC GTG ATC ACC CAG TAC CAG TCC CAC ACT CTG    240
Gly Ile Arg Arg Met Gly Val Ile Thr Gln Tyr Gln Ser His Thr Leu
65                  70                  75                  80

GTG CAG CAC ATT CAG CGC GGC TGG TCA TTC TTC AAT GAA GAA ATG AAC    288
Val Gln His Ile Gln Arg Gly Trp Ser Phe Phe Asn Glu Glu Met Asn
                85                  90                  95

GAG TTT GTC GAT CTG CTG CCA GCA CAG CAG AGA ATG AAA GGG GAA AAC    336
Glu Phe Val Asp Leu Leu Pro Ala Gln Gln Arg Met Lys Gly Glu Asn
            100                 105                 110

TGG TAT CGC GGC ACC GCA GAT GCG GTC ACC CAA AAC CTC GAC ATT ATC    384
Trp Tyr Arg Gly Thr Ala Asp Ala Val Thr Gln Asn Leu Asp Ile Ile
        115                 120                 125

CGT CGT TAT AAA GCG GAA TAC GTG GTG ATC CTG GCG GGC GAC CAT ATC    432
Arg Arg Tyr Lys Ala Glu Tyr Val Val Ile Leu Ala Gly Asp His Ile
    130                 135                 140

TAC AAG CAA GAC TAC TCG CGT ATG CTT ATC GAT CAC GTC GAA AAA GGT    480
Tyr Lys Gln Asp Tyr Ser Arg Met Leu Ile Asp His Val Glu Lys Gly
145                 150                 155                 160

GTA CGT TGT ACC GTT GTT TGT ATG CCA GTA CCG ATT GAA GAA GCC TCC    528
Val Arg Cys Thr Val Val Cys Met Pro Val Pro Ile Glu Glu Ala Ser
                165                 170                 175

GCA TTT GGC GTT ATG GCG GTT GAT GAG AAC GAT AAA ACT ATC GAA TTC    576
Ala Phe Gly Val Met Ala Val Asp Glu Asn Asp Lys Thr Ile Glu Phe
            180                 185                 190

GTG GAA AAA CCT GCT AAC CCG CCG TCA ATG CCG AAC GAT CCG AGC AAA    624
Val Glu Lys Pro Ala Asn Pro Pro Ser Met Pro Asn Asp Pro Ser Lys
```

```
                195                 200                 205

TCT CTG GCG AGT ATG GGT ATC TAC GTC TTT GAC GCC GAC TAT CTG TAT   672
Ser Leu Ala Ser Met Gly Ile Tyr Val Phe Asp Ala Asp Tyr Leu Tyr
    210                 215                 220

GAA CTG CTG GAA GAA GAC GAT CGC GAT GAG AAC TCC AGC CAC GAC TTT   720
Glu Leu Leu Glu Glu Asp Asp Arg Asp Glu Asn Ser Ser His Asp Phe
225                 230                 235                 240

GGC AAA GAT TTG ATT CCC AAG ATC ACC GAA GCC GGT CTG GCC TAT GCG   768
Gly Lys Asp Leu Ile Pro Lys Ile Thr Glu Ala Gly Leu Ala Tyr Ala
                245                 250                 255

CAC CCG TTC CCG CTC TCT TGC GTA CAA TCC GAC CCG GAT GCC GAG CCG   816
His Pro Phe Pro Leu Ser Cys Val Gln Ser Asp Pro Asp Ala Glu Pro
            260                 265                 270

TAC TGG CGC GAT GTG GGT ACG CTG GAA GCT TAC TGG AAA GCG AAC CTC   864
Tyr Trp Arg Asp Val Gly Thr Leu Glu Ala Tyr Trp Lys Ala Asn Leu
        275                 280                 285

GAT CTG GCC TCT GTG GTG CCG GAG CTG GAT ATG TAC GAT CGC AAT TGG   912
Asp Leu Ala Ser Val Val Pro Glu Leu Asp Met Tyr Asp Arg Asn Trp
    290                 295                 300

CCA ATT CGC ACC TAC AAT GAA TCA TTA CCG CCA GCG AAA TTC GTG CAG   960
Pro Ile Arg Thr Tyr Asn Glu Ser Leu Pro Pro Ala Lys Phe Val Gln
305                 310                 315                 320

GAT CGC TCC GGT AGC CAC GGG ATG ACC CTT AAC TCA CTG GTT TCC GGC  1008
Asp Arg Ser Gly Ser His Gly Met Thr Leu Asn Ser Leu Val Ser Gly
                325                 330                 335

GGT TGT GTG ATC TCC GGT TCG GTG GTG GTG CAG TCC GTT CTG TTC TCG  1056
Gly Cys Val Ile Ser Gly Ser Val Val Val Gln Ser Val Leu Phe Ser
            340                 345                 350

CGC GTT CGC GTG AAT TCA TTC TGC AAC ATT GAT TCC GCC GTA TTG TTA  1104
Arg Val Arg Val Asn Ser Phe Cys Asn Ile Asp Ser Ala Val Leu Leu
        355                 360                 365

CCG GAA GTA TGG GTA GGT CGC TCG TGC CGT CTG CGC CGC TGC GTC ATC  1152
Pro Glu Val Trp Val Gly Arg Ser Cys Arg Leu Arg Arg Cys Val Ile
    370                 375                 380

GAT CGT GCT TGT GTT ATT CCG GAA GGC ATG GTG ATT GGT GAA AAC GCA  1200
Asp Arg Ala Cys Val Ile Pro Glu Gly Met Val Ile Gly Glu Asn Ala
385                 390                 395                 400

GAG GAA GAT GCA CGT CGT TTC TAT CGT TCA GAA GAA GGC ATC GTG CTG  1248
Glu Glu Asp Ala Arg Arg Phe Tyr Arg Ser Glu Glu Gly Ile Val Leu
                405                 410                 415

GTA ACG CGC GAA ATG CTA CGG AAG TTA GGG CAT AAA CAG GAG CGA TAA  1296
Val Thr Arg Glu Met Leu Arg Lys Leu Gly His Lys Gln Glu Arg
            420                 425                 430
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 431 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 431

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Val Ser Leu Glu Lys Asn Asp His Leu Met Leu Ala Arg Gln Leu
 1               5                  10                  15

Pro Leu Lys Ser Val Ala Leu Ile Leu Ala Gly Gly Arg Gly Thr Arg
            20                  25                  30

Leu Lys Asp Leu Thr Asn Lys Arg Ala Lys Pro Ala Val His Phe Gly
        35                  40                  45

Gly Lys Phe Arg Ile Ile Asp Phe Ala Leu Ser Asn Cys Ile Asn Ser
    50                  55                  60

Gly Ile Arg Arg Met Gly Val Ile Thr Gln Tyr Gln Ser His Thr Leu
65                  70                  75                  80

Val Gln His Ile Gln Arg Gly Trp Ser Phe Phe Asn Glu Glu Met Asn
                85                  90                  95

Glu Phe Val Asp Leu Leu Pro Ala Gln Gln Arg Met Lys Gly Glu Asn
            100                 105                 110

Trp Tyr Arg Gly Thr Ala Asp Ala Val Thr Gln Asn Leu Asp Ile Ile
        115                 120                 125

Arg Arg Tyr Lys Ala Glu Tyr Val Val Ile Leu Ala Gly Asp His Ile
    130                 135                 140

Tyr Lys Gln Asp Tyr Ser Arg Met Leu Ile Asp His Val Glu Lys Gly
145                 150                 155                 160

Val Arg Cys Thr Val Val Cys Met Pro Val Pro Ile Glu Glu Ala Ser
                165                 170                 175

Ala Phe Gly Val Met Ala Val Asp Glu Asn Asp Lys Thr Ile Glu Phe
            180                 185                 190

Val Glu Lys Pro Ala Asn Pro Pro Ser Met Pro Asn Asp Pro Ser Lys
```

195 200 205

Ser Leu Ala Ser Met Gly Ile Tyr Val Phe Asp Ala Asp Tyr Leu Tyr
210 215 220

Glu Leu Leu Glu Glu Asp Asp Arg Asp Glu Asn Ser Ser His Asp Phe
225 230 235 240

Gly Lys Asp Leu Ile Pro Lys Ile Thr Glu Ala Gly Leu Ala Tyr Ala
245 250 255

His Pro Phe Pro Leu Ser Cys Val Gln Ser Asp Pro Asp Ala Glu Pro
260 265 270

Tyr Trp Arg Asp Val Gly Thr Leu Glu Ala Tyr Trp Lys Ala Asn Leu
275 280 285

Asp Leu Ala Ser Val Val Pro Glu Leu Asp Met Tyr Asp Arg Asn Trp
290 295 300

Pro Ile Arg Thr Tyr Asn Glu Ser Leu Pro Pro Ala Lys Phe Val Gln
305 310 315 320

Asp Arg Ser Gly Ser His Gly Met Thr Leu Asn Ser Leu Val Ser Gly
325 330 335

Gly Cys Val Ile Ser Gly Ser Val Val Val Gln Ser Val Leu Phe Ser
340 345 350

Arg Val Arg Val Asn Ser Phe Cys Asn Ile Asp Ser Ala Val Leu Leu
355 360 365

Pro Glu Val Trp Val Gly Arg Ser Cys Arg Leu Arg Arg Cys Val Ile
370 375 380

Asp Arg Ala Cys Val Ile Pro Glu Gly Met Val Ile Gly Glu Asn Ala
385 390 395 400

Glu Glu Asp Ala Arg Arg Phe Tyr Arg Ser Glu Glu Gly Ile Val Leu
405 410 415

Val Thr Arg Glu Met Leu Arg Lys Leu Gly His Lys Gln Glu Arg
420 425 430

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1296 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..1293

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 TO 1296

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG GTT AGT TTA GAG AAG AAC GAT CAC TTA ATG TTG GCG CGC CAG CTG     48
Met Val Ser Leu Glu Lys Asn Asp His Leu Met Leu Ala Arg Gln Leu
  1               5                  10                  15

CCA TTG AAA TCT GTT GCC CTG ATA CTG GCG GGA GGA CGT GGT ACC CGC     96
Pro Leu Lys Ser Val Ala Leu Ile Leu Ala Gly Gly Arg Gly Thr Arg
             20                  25                  30

CTG AAG GAT TTA ACC AAT AAG CGA GCA AAA CCG GCC GTA CAC TTC GGC    144
Leu Lys Asp Leu Thr Asn Lys Arg Ala Lys Pro Ala Val His Phe Gly
         35                  40                  45

GGT AAG TTC CGC ATT ATC GAC TTT GCG CTG TCT AAC TGC ATC AAC TCC    192
Gly Lys Phe Arg Ile Ile Asp Phe Ala Leu Ser Asn Cys Ile Asn Ser
     50                  55                  60

GGG ATC CGT CGT ATG GGC GTG ATC ACC CAG TAC CAG TCC CAC ACT CTG    240
Gly Ile Arg Arg Met Gly Val Ile Thr Gln Tyr Gln Ser His Thr Leu
 65                  70                  75                  80

GTG CAG CAC ATT CAG CGC GGC TGG TCA TTC TTC AAT GAA GAA ATG AAC    288
Val Gln His Ile Gln Arg Gly Trp Ser Phe Phe Asn Glu Glu Met Asn
                 85                  90                  95

GAG TTT GTC GAT CTG CTG CCA GCA CAG CAG AGA ATG AAA GGG GAA AAC    336
Glu Phe Val Asp Leu Leu Pro Ala Gln Gln Arg Met Lys Gly Glu Asn
            100                 105                 110

TGG TAT CGC GGC ACC GCA GAT GCG GTC ACC CAA AAC CTC GAC ATT ATC    384
Trp Tyr Arg Gly Thr Ala Asp Ala Val Thr Gln Asn Leu Asp Ile Ile
        115                 120                 125

CGT CGT TAT AAA GCG GAA TAC GTG GTG ATC CTG GCG GGC GAC CAT ATC    432
Arg Arg Tyr Lys Ala Glu Tyr Val Val Ile Leu Ala Gly Asp His Ile
    130                 135                 140

TAC AAG CAA GAC TAC TCG CGT ATG CTT ATC GAT CAC GTC GAA AAA GGT    480
Tyr Lys Gln Asp Tyr Ser Arg Met Leu Ile Asp His Val Glu Lys Gly
145                 150                 155                 160

GTA CGT TGT ACC GTT GTT TGT ATG CCA GTA CCG ATT GAA GAA GCC TCC    528
Val Arg Cys Thr Val Val Cys Met Pro Val Pro Ile Glu Glu Ala Ser
                165                 170                 175
```

```
GCA TTT GGC GTT ATG GCG GTT GAT GAG AAC GAT AAA ACT ATC GAA TTC   576
Ala Phe Gly Val Met Ala Val Asp Glu Asn Asp Lys Thr Ile Glu Phe
            180                 185                 190

GTG GAA AAA CCT GCT AAC CCG CCG TCA ATG CCG AAC GAT CCG AGC AAA   624
Val Glu Lys Pro Ala Asn Pro Pro Ser Met Pro Asn Asp Pro Ser Lys
        195                 200                 205

TCT CTG GCG AGT ATG GGT ATC TAC GTC TTT GAC GCC GAC TAT CTG TAT   672
Ser Leu Ala Ser Met Gly Ile Tyr Val Phe Asp Ala Asp Tyr Leu Tyr
    210                 215                 220

GAA CTG CTG GAA GAA GAC GAT CGC GAT GAG AAC TCC AGC CAC GAC TTT   720
Glu Leu Leu Glu Glu Asp Asp Arg Asp Glu Asn Ser Ser His Asp Phe
225                 230                 235                 240

GGC AAA GAT TTG ATT CCC AAG ATC ACC GAA GCC GGT CTG GCC TAT GCG   768
Gly Lys Asp Leu Ile Pro Lys Ile Thr Glu Ala Gly Leu Ala Tyr Ala
                245                 250                 255

CAC CCG TTC CCG CTC TCT TGC GTA CAA TCC GAC CCG GAT GCC GAG CCG   816
His Pro Phe Pro Leu Ser Cys Val Gln Ser Asp Pro Asp Ala Glu Pro
            260                 265                 270

TAC TGG CGC GAT GTG GGT ACG CTG GAA GCT TAC TGG AAA GCG AAC CTC   864
Tyr Trp Arg Asp Val Gly Thr Leu Glu Ala Tyr Trp Lys Ala Asn Leu
        275                 280                 285

GAT CTG GCC TCT GTG GTG CCG GAG CTG GAT ATG TAC GAT CGC AAT TGG   912
Asp Leu Ala Ser Val Val Pro Glu Leu Asp Met Tyr Asp Arg Asn Trp
    290                 295                 300

CCA ATT CGC ACC TAC AAT GAA TCA TTA CCG CCA GCG AAA TTC GTG CAG   960
Pro Ile Arg Thr Tyr Asn Glu Ser Leu Pro Pro Ala Lys Phe Val Gln
305                 310                 315                 320

GAT CGC TCC GGT AGC CAC GGG ATG ACC CTT AAC TCA CTG GTT TCC GAC  1008
Asp Arg Ser Gly Ser His Gly Met Thr Leu Asn Ser Leu Val Ser Asp
                325                 330                 335

GGT TGT GTG ATC TCC GGT TCG GTG GTG GTG CAG TCC GTT CTG TTC TCG  1056
Gly Cys Val Ile Ser Gly Ser Val Val Val Gln Ser Val Leu Phe Ser
            340                 345                 350

CGC GTT CGC GTG AAT TCA TTC TGC AAC ATT GAT TCC GCC GTA TTG TTA  1104
Arg Val Arg Val Asn Ser Phe Cys Asn Ile Asp Ser Ala Val Leu Leu
        355                 360                 365

CCG GAA GTA TGG GTA GGT CGC TCG TGC CGT CTG CGC CGC TGC GTC ATC  1152
Pro Glu Val Trp Val Gly Arg Ser Cys Arg Leu Arg Arg Cys Val Ile
    370                 375                 380

GAT CGT GCT TGT GTT ATT CCG GAA GGC ATG GTG ATT GGT GAA AAC GCA  1200
Asp Arg Ala Cys Val Ile Pro Glu Gly Met Val Ile Gly Glu Asn Ala
385                 390                 395                 400

GAG GAA GAT GCA CGT CGT TTC TAT CGT TCA GAA GAA GGC ATC GTG CTG  1248
```

```
Glu Glu Asp Ala Arg Arg Phe Tyr Arg Ser Glu Glu Gly Ile Val Leu
                405                 410                 415

GTA ACG CGC GAA ATG CTA CGG AAG TTA GGG CAT AAA CAG GAG CGA TAA  1296
Val Thr Arg Glu Met Leu Arg Lys Leu Gly His Lys Gln Glu Arg
            420                 425                 430
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 431 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 4: FROM 1 TO 431

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Val Ser Leu Glu Lys Asn Asp His Leu Met Leu Ala Arg Gln Leu
 1               5                  10                  15

Pro Leu Lys Ser Val Ala Leu Ile Leu Ala Gly Gly Arg Gly Thr Arg
            20                  25                  30

Leu Lys Asp Leu Thr Asn Lys Arg Ala Lys Pro Ala Val His Phe Gly
        35                  40                  45

Gly Lys Phe Arg Ile Ile Asp Phe Ala Leu Ser Asn Cys Ile Asn Ser
    50                  55                  60

Gly Ile Arg Arg Met Gly Val Ile Thr Gln Tyr Gln Ser His Thr Leu
65                  70                  75                  80

Val Gln His Ile Gln Arg Gly Trp Ser Phe Phe Asn Glu Glu Met Asn
                85                  90                  95

Glu Phe Val Asp Leu Leu Pro Ala Gln Gln Arg Met Lys Gly Glu Asn
            100                 105                 110

Trp Tyr Arg Gly Thr Ala Asp Ala Val Thr Gln Asn Leu Asp Ile Ile
        115                 120                 125

Arg Arg Tyr Lys Ala Glu Tyr Val Val Ile Leu Ala Gly Asp His Ile
    130                 135                 140

Tyr Lys Gln Asp Tyr Ser Arg Met Leu Ile Asp His Val Glu Lys Gly
145                 150                 155                 160

Val Arg Cys Thr Val Val Cys Met Pro Val Pro Ile Glu Glu Ala Ser
```

```
                165                 170                 175

Ala Phe Gly Val Met Ala Val Asp Glu Asn Asp Lys Thr Ile Glu Phe
            180                 185                 190

Val Glu Lys Pro Ala Asn Pro Pro Ser Met Pro Asn Asp Pro Ser Lys
        195                 200                 205

Ser Leu Ala Ser Met Gly Ile Tyr Val Phe Asp Ala Asp Tyr Leu Tyr
    210                 215                 220

Glu Leu Leu Glu Glu Asp Asp Arg Asp Glu Asn Ser Ser His Asp Phe
225                 230                 235                 240

Gly Lys Asp Leu Ile Pro Lys Ile Thr Glu Ala Gly Leu Ala Tyr Ala
                245                 250                 255

His Pro Phe Pro Leu Ser Cys Val Gln Ser Asp Pro Asp Ala Glu Pro
            260                 265                 270

Tyr Trp Arg Asp Val Gly Thr Leu Glu Ala Tyr Trp Lys Ala Asn Leu
        275                 280                 285

Asp Leu Ala Ser Val Val Pro Glu Leu Asp Met Tyr Asp Arg Asn Trp
    290                 295                 300

Pro Ile Arg Thr Tyr Asn Glu Ser Leu Pro Pro Ala Lys Phe Val Gln
305                 310                 315                 320

Asp Arg Ser Gly Ser His Gly Met Thr Leu Asn Ser Leu Val Ser Asp
                325                 330                 335

Gly Cys Val Ile Ser Gly Ser Val Val Val Gln Ser Val Leu Phe Ser
            340                 345                 350

Arg Val Arg Val Asn Ser Phe Cys Asn Ile Asp Ser Ala Val Leu Leu
        355                 360                 365

Pro Glu Val Trp Val Gly Arg Ser Cys Arg Leu Arg Arg Cys Val Ile
    370                 375                 380

Asp Arg Ala Cys Val Ile Pro Glu Gly Met Val Ile Gly Glu Asn Ala
385                 390                 395                 400

Glu Glu Asp Ala Arg Arg Phe Tyr Arg Ser Glu Glu Gly Ile Val Leu
                405                 410                 415

Val Thr Arg Glu Met Leu Arg Lys Leu Gly His Lys Gln Glu Arg
            420                 425                 430
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 88..354

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 355

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
AAGCTTGTTC TCATTGTTGT TATCATTATA TATAGATGAC CAAAGCACTA GACCAAACCT   60

CAGTCACACA AAGAGTAAAG AAGAACA ATG GCT TCC TCT ATG CTC TCT TCC       111
                              Met Ala Ser Ser Met Leu Ser Ser
                                1               5

GCT ACT ATG GTT GCC TCT CCG GCT CAG GCC ACT ATG GTC GCT CCT TTC     159
Ala Thr Met Val Ala Ser Pro Ala Gln Ala Thr Met Val Ala Pro Phe
     10                  15                  20

AAC GGA CTT AAG TCC TCC GCT GCC TTC CCA GCC ACC CGC AAG GCT AAC     207
Asn Gly Leu Lys Ser Ser Ala Ala Phe Pro Ala Thr Arg Lys Ala Asn
 25                  30                  35                  40

AAC GAC ATT ACT TCC ATC ACA AGC AAC GGC GGA AGA GTT AAC TGC ATG     255
Asn Asp Ile Thr Ser Ile Thr Ser Asn Gly Gly Arg Val Asn Cys Met
                    45                  50                  55

CAG GTG TGG CCT CCG ATT GGA AAG AAG AAG TTT GAG ACT CTC TCT TAC     303
Gln Val Trp Pro Pro Ile Gly Lys Lys Lys Phe Glu Thr Leu Ser Tyr
               60                  65                  70

CTT CCT GAC CTT ACC GAT TCC GGT GGT CGC GTC AAC TGC ATG CAG GCC     351
Leu Pro Asp Leu Thr Asp Ser Gly Gly Arg Val Asn Cys Met Gln Ala
          75                  80                  85

ATG G                                                               355
Met
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 89 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 89

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala
 1               5                  10                  15

Gln Ala Thr Met Val Ala Pro Phe Asn Gly Leu Lys Ser Ser Ala Ala
            20                  25                  30

Phe Pro Ala Thr Arg Lys Ala Asn Asn Asp Ile Thr Ser Ile Thr Ser
        35                  40                  45

Asn Gly Gly Arg Val Asn Cys Met Gln Val Trp Pro Pro Ile Gly Lys
    50                  55                  60

Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Asp Leu Thr Asp Ser Gly
65                  70                  75                  80

Gly Arg Val Asn Cys Met Gln Ala Met
                85
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1575 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..1565

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 1575

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CC ATG GCG GCT TCC ATT GGA GCC TTA AAA TCT TCA CCT TCT TCT AAC      47
   Met Ala Ala Ser Ile Gly Ala Leu Lys Ser Ser Pro Ser Ser Asn
    1               5                  10                  15
```

```
AAT TGC ATC AAT GAG AGA AGA AAT GAT TCT ACA CGT GCT GTA TCC AGC      95
Asn Cys Ile Asn Glu Arg Arg Asn Asp Ser Thr Arg Ala Val Ser Ser
                20                  25                  30

AGA AAT CTC TCA TTT TCG TCT TCT CAT CTC GCC GGA GAC AAG TTG ATG     143
Arg Asn Leu Ser Phe Ser Ser Ser His Leu Ala Gly Asp Lys Leu Met
            35                  40                  45

CCT GTA TCG TCC TTA CGT TCC CAA GGA GTC CGA TTC AAT GTG AGA AGA     191
Pro Val Ser Ser Leu Arg Ser Gln Gly Val Arg Phe Asn Val Arg Arg
        50                  55                  60

AGT CCA ATG ATT GTG TCG CCA AAG GCT GTT TCT GAT TCG CAG AAT TCA     239
Ser Pro Met Ile Val Ser Pro Lys Ala Val Ser Asp Ser Gln Asn Ser
    65                  70                  75

CAG ACA TGT CTA GAC CCA GAT GCT AGC CGG AGT GTT TTG GGA ATT ATT     287
Gln Thr Cys Leu Asp Pro Asp Ala Ser Arg Ser Val Leu Gly Ile Ile
80                  85                  90                  95

CTT GGA GGT GGA GCT GGG ACC CGA CTT TAT CCT CTA ACT AAA AAA AGA     335
Leu Gly Gly Gly Ala Gly Thr Arg Leu Tyr Pro Leu Thr Lys Lys Arg
                100                 105                 110

GCA AAG CCA GCT GTT CCA CTT GGA GCA AAT TAT CGT CTG ATT GAC ATT     383
Ala Lys Pro Ala Val Pro Leu Gly Ala Asn Tyr Arg Leu Ile Asp Ile
            115                 120                 125

CCT GTA AGC AAC TGC TTG AAC AGT AAT ATA TCC AAG ATT TAT GTT CTC     431
Pro Val Ser Asn Cys Leu Asn Ser Asn Ile Ser Lys Ile Tyr Val Leu
        130                 135                 140

ACA CAA TTC AAC TCT GCC TCT CTG AAT CGC CAC CTT TCA CGA GCA TAT     479
Thr Gln Phe Asn Ser Ala Ser Leu Asn Arg His Leu Ser Arg Ala Tyr
    145                 150                 155

GCT AGC AAC ATG GGA GGA TAC AAA AAC GAG GGC TTT GTG GAA GTT CTT     527
Ala Ser Asn Met Gly Gly Tyr Lys Asn Glu Gly Phe Val Glu Val Leu
160                 165                 170                 175

GCT GCT CAA CAA AGT CCA GAG AAC CCC GAT TGG TTC CAG GGC ACG GCT     575
Ala Ala Gln Gln Ser Pro Glu Asn Pro Asp Trp Phe Gln Gly Thr Ala
                180                 185                 190

GAT GCT GTC AGA CAA TAT CTG TGG TTG TTT GAG GAG CAT ACT GTT CTT     623
Asp Ala Val Arg Gln Tyr Leu Trp Leu Phe Glu Glu His Thr Val Leu
            195                 200                 205

GAA TAC CTT ATA CTT GCT GGA GAT CAT CTG TAT CGA ATG GAT TAT GAA     671
Glu Tyr Leu Ile Leu Ala Gly Asp His Leu Tyr Arg Met Asp Tyr Glu
        210                 215                 220

AAG TTT ATT CAA GCC CAC AGA GAA ACA GAT GCT GAT ATT ACC GTT GCC     719
Lys Phe Ile Gln Ala His Arg Glu Thr Asp Ala Asp Ile Thr Val Ala
    225                 230                 235

GCA CTG CCA ATG GAC GAG AAG CGT GCC ACT GCA TTC GGT CTC ATG AAG     767
```

```
Ala Leu Pro Met Asp Glu Lys Arg Ala Thr Ala Phe Gly Leu Met Lys
240                 245                 250                 255

ATT GAC GAA GAA GGA CGC ATT ATT GAA TTT GCA GAG AAA CCG CAA GGA   815
Ile Asp Glu Glu Gly Arg Ile Ile Glu Phe Ala Glu Lys Pro Gln Gly
                260                 265                 270

GAG CAA TTG CAA GCA ATG AAA GTG GAT ACT ACC ATT TTA GGT CTT GAT   863
Glu Gln Leu Gln Ala Met Lys Val Asp Thr Thr Ile Leu Gly Leu Asp
            275                 280                 285

GAC AAG AGA GCT AAA GAA ATG CCT TTC ATT GCC AGT ATG GGT ATA TAT   911
Asp Lys Arg Ala Lys Glu Met Pro Phe Ile Ala Ser Met Gly Ile Tyr
        290                 295                 300

GTC ATT AGC AAA GAC GTG ATG TTA AAC CTA CTT CGT GAC AAG TTC CCT   959
Val Ile Ser Lys Asp Val Met Leu Asn Leu Leu Arg Asp Lys Phe Pro
    305                 310                 315

GGG GCC AAT GAT TTT GGT AGT GAA GTT ATT CCT GGT GCA ACT TCA CTT  1007
Gly Ala Asn Asp Phe Gly Ser Glu Val Ile Pro Gly Ala Thr Ser Leu
320                 325                 330                 335

GGG ATG AGA GTG CAA GCT TAT TTA TAT GAT GGG TAC TGG GAA GAT ATT  1055
Gly Met Arg Val Gln Ala Tyr Leu Tyr Asp Gly Tyr Trp Glu Asp Ile
                340                 345                 350

GGT ACC ATT GAA GCT TTC TAC AAT GCC AAT TTG GGC ATT ACA AAA AAG  1103
Gly Thr Ile Glu Ala Phe Tyr Asn Ala Asn Leu Gly Ile Thr Lys Lys
            355                 360                 365

CCG GTG CCA GAT TTT AGC TTT TAC GAC CGA TCA GCC CCA ATC TAC ACC  1151
Pro Val Pro Asp Phe Ser Phe Tyr Asp Arg Ser Ala Pro Ile Tyr Thr
        370                 375                 380

CAA CCT CGA TAT CTA CCA CCA TCA AAA ATG CTT GAT GCT GAT GTC ACA  1199
Gln Pro Arg Tyr Leu Pro Pro Ser Lys Met Leu Asp Ala Asp Val Thr
    385                 390                 395

GAT AGT GTC ATT GGT GAA GGT TGT GTG ATC AAG AAC TGT AAG ATT CAT  1247
Asp Ser Val Ile Gly Glu Gly Cys Val Ile Lys Asn Cys Lys Ile His
400                 405                 410                 415

CAT TCC GTG GTT GGA CTC AGA TCA TGC ATA TCA GAG GGA GCA ATT ATA  1295
His Ser Val Val Gly Leu Arg Ser Cys Ile Ser Glu Gly Ala Ile Ile
                420                 425                 430

GAA GAC TCA CTT TTG ATG GGG GCA GAT TAC TAT GAG ACT GAT GCT GAC  1343
Glu Asp Ser Leu Leu Met Gly Ala Asp Tyr Tyr Glu Thr Asp Ala Asp
            435                 440                 445

AGG AAG TTG CTG GCT GCA AAG GGC AGT GTC CCA ATT GGC ATC GGC AAG  1391
Arg Lys Leu Leu Ala Ala Lys Gly Ser Val Pro Ile Gly Ile Gly Lys
        450                 455                 460

AAT TGT CAC ATT AAA AGA GCC ATT ATC GAC AAG AAT GCC CGT ATA GGG  1439
Asn Cys His Ile Lys Arg Ala Ile Ile Asp Lys Asn Ala Arg Ile Gly
```

```
          465                 470                 475

GAC AAT GTG AAG ATC ATT AAC AAA GAC AAC GTT CAA GAA GCG GCT AGG  1487
Asp Asn Val Lys Ile Ile Asn Lys Asp Asn Val Gln Glu Ala Ala Arg
480                 485                 490                 495

GAA ACA GAT GGA TAC TTC ATC AAG AGT GGG ATT GTC ACC GTC ATC AAG  1535
Glu Thr Asp Gly Tyr Phe Ile Lys Ser Gly Ile Val Thr Val Ile Lys
                500                 505                 510

GAT GCT TTG ATT CCA AGT GGA ATC ATC ATC TGATGAGCTC               1575
Asp Ala Leu Ile Pro Ser Gly Ile Ile Ile
            515                 520
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 521 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 8: FROM 1 TO 521

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Ala Ala Ser Ile Gly Ala Leu Lys Ser Ser Pro Ser Ser Asn Asn
 1               5                  10                  15

Cys Ile Asn Glu Arg Arg Asn Asp Ser Thr Arg Ala Val Ser Ser Arg
            20                  25                  30

Asn Leu Ser Phe Ser Ser Ser His Leu Ala Gly Asp Lys Leu Met Pro
        35                  40                  45

Val Ser Ser Leu Arg Ser Gln Gly Val Arg Phe Asn Val Arg Arg Ser
    50                  55                  60

Pro Met Ile Val Ser Pro Lys Ala Val Ser Asp Ser Gln Asn Ser Gln
65                  70                  75                  80

Thr Cys Leu Asp Pro Asp Ala Ser Arg Ser Val Leu Gly Ile Ile Leu
                85                  90                  95

Gly Gly Gly Ala Gly Thr Arg Leu Tyr Pro Leu Thr Lys Lys Arg Ala
            100                 105                 110

Lys Pro Ala Val Pro Leu Gly Ala Asn Tyr Arg Leu Ile Asp Ile Pro
        115                 120                 125
```

```
Val Ser Asn Cys Leu Asn Ser Asn Ile Ser Lys Ile Tyr Val Leu Thr
    130                 135                 140

Gln Phe Asn Ser Ala Ser Leu Asn Arg His Leu Ser Arg Ala Tyr Ala
145                 150                 155                 160

Ser Asn Met Gly Gly Tyr Lys Asn Glu Gly Phe Val Glu Val Leu Ala
                165                 170                 175

Ala Gln Gln Ser Pro Glu Asn Pro Asp Trp Phe Gln Gly Thr Ala Asp
            180                 185                 190

Ala Val Arg Gln Tyr Leu Trp Leu Phe Glu Glu His Thr Val Leu Glu
        195                 200                 205

Tyr Leu Ile Leu Ala Gly Asp His Leu Tyr Arg Met Asp Tyr Glu Lys
    210                 215                 220

Phe Ile Gln Ala His Arg Glu Thr Asp Ala Asp Ile Thr Val Ala Ala
225                 230                 235                 240

Leu Pro Met Asp Glu Lys Arg Ala Thr Ala Phe Gly Leu Met Lys Ile
                245                 250                 255

Asp Glu Glu Gly Arg Ile Ile Glu Phe Ala Glu Lys Pro Gln Gly Glu
            260                 265                 270

Gln Leu Gln Ala Met Lys Val Asp Thr Thr Ile Leu Gly Leu Asp Asp
        275                 280                 285

Lys Arg Ala Lys Glu Met Pro Phe Ile Ala Ser Met Gly Ile Tyr Val
    290                 295                 300

Ile Ser Lys Asp Val Met Leu Asn Leu Leu Arg Asp Lys Phe Pro Gly
305                 310                 315                 320

Ala Asn Asp Phe Gly Ser Glu Val Ile Pro Gly Ala Thr Ser Leu Gly
                325                 330                 335

Met Arg Val Gln Ala Tyr Leu Tyr Asp Gly Tyr Trp Glu Asp Ile Gly
            340                 345                 350

Thr Ile Glu Ala Phe Tyr Asn Ala Asn Leu Gly Ile Thr Lys Lys Pro
        355                 360                 365

Val Pro Asp Phe Ser Phe Tyr Asp Arg Ser Ala Pro Ile Tyr Thr Gln
    370                 375                 380

Pro Arg Tyr Leu Pro Pro Ser Lys Met Leu Asp Ala Asp Val Thr Asp
385                 390                 395                 400

Ser Val Ile Gly Glu Gly Cys Val Ile Lys Asn Cys Lys Ile His His
                405                 410                 415

Ser Val Val Gly Leu Arg Ser Cys Ile Ser Glu Gly Ala Ile Ile Glu
            420                 425                 430
```

```
Asp Ser Leu Leu Met Gly Ala Asp Tyr Tyr Glu Thr Asp Ala Asp Arg
        435                 440                 445

Lys Leu Leu Ala Ala Lys Gly Ser Val Pro Ile Gly Ile Gly Lys Asn
    450                 455                 460

Cys His Ile Lys Arg Ala Ile Ile Asp Lys Asn Ala Arg Ile Gly Asp
465                 470                 475                 480

Asn Val Lys Ile Ile Asn Lys Asp Asn Val Gln Glu Ala Ala Arg Glu
                485                 490                 495

Thr Asp Gly Tyr Phe Ile Lys Ser Gly Ile Val Thr Val Ile Lys Asp
            500                 505                 510

Ala Leu Ile Pro Ser Gly Ile Ile Ile
        515                 520
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1519 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1410

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 9: FROM 1 TO 1519

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
AAC AAG ATC AAA CCT GGG GTT GCT TAC TCT GTG ATC ACT ACT GAA AAT    48
Asn Lys Ile Lys Pro Gly Val Ala Tyr Ser Val Ile Thr Thr Glu Asn
 1               5                  10                  15

GAC ACA CAG ACT GTG TTC GTA GAT ATG CCA CGT CTT GAG AGA CGC CGG    96
Asp Thr Gln Thr Val Phe Val Asp Met Pro Arg Leu Glu Arg Arg Arg
            20                  25                  30

GCA AAT CCA AAG GAT GTG GCT GCA GTC ATA CTG GGA GGA GGA GAA GGG   144
Ala Asn Pro Lys Asp Val Ala Ala Val Ile Leu Gly Gly Gly Glu Gly
        35                  40                  45

ACC AAG TTA TTC CCA CTT ACA AGT AGA ACT GCA ACC CCT GCT GTT CCG   192
Thr Lys Leu Phe Pro Leu Thr Ser Arg Thr Ala Thr Pro Ala Val Pro
```

```
     50                  55                  60

GTT GGA GGA TGC TAC AGG CTA ATA GAC ATC CCA ATG AGC AAC TGT ATC   240
Val Gly Gly Cys Tyr Arg Leu Ile Asp Ile Pro Met Ser Asn Cys Ile
 65                  70                  75                  80

AAC AGT GCT ATT AAC AAG ATT TTT GTG CTG ACA CAG TAC AAT TCT GCT   288
Asn Ser Ala Ile Asn Lys Ile Phe Val Leu Thr Gln Tyr Asn Ser Ala
                 85                  90                  95

CCC CTG AAT CGT CAC ATT GCT CGA ACA TAT TTT GGC AAT GGT GTG AGC   336
Pro Leu Asn Arg His Ile Ala Arg Thr Tyr Phe Gly Asn Gly Val Ser
            100                 105                 110

TTT GGA GAT GGA TTT GTC GAG GTA CTA GCT GCA ACT CAG ACA CCC GGG   384
Phe Gly Asp Gly Phe Val Glu Val Leu Ala Ala Thr Gln Thr Pro Gly
        115                 120                 125

GAA GCA GGA AAA AAA TGG TTT CAA GGA ACA GCA GAT GCT GTT AGA AAA   432
Glu Ala Gly Lys Lys Trp Phe Gln Gly Thr Ala Asp Ala Val Arg Lys
    130                 135                 140

TTT ATA TGG GTT TTT GAG GAC GCT AAG AAC AAG AAT ATT GAA AAT ATC   480
Phe Ile Trp Val Phe Glu Asp Ala Lys Asn Lys Asn Ile Glu Asn Ile
145                 150                 155                 160

GTT GTA CTA TCT GGG GAT CAT CTT TAT AGG ATG GAT TAT ATG GAG TTG   528
Val Val Leu Ser Gly Asp His Leu Tyr Arg Met Asp Tyr Met Glu Leu
                165                 170                 175

GTG CAG AAC CAT ATT GAC AGG AAT GCT GAT ATT ACT CTT TCA TGT GCA   576
Val Gln Asn His Ile Asp Arg Asn Ala Asp Ile Thr Leu Ser Cys Ala
            180                 185                 190

CCA GCT GAG GAC AGC CGA GCA TCA GAT TTT GGG CTG GTC AAG ATT GAC   624
Pro Ala Glu Asp Ser Arg Ala Ser Asp Phe Gly Leu Val Lys Ile Asp
        195                 200                 205

AGC AGA GGC AGA GTA GTC CAG TTT GCT GAA AAA CCA AAA GGT TTT GAT   672
Ser Arg Gly Arg Val Val Gln Phe Ala Glu Lys Pro Lys Gly Phe Asp
    210                 215                 220

CTT AAA GCA ATG CAA GTA GAT ACT ACT CTT GTT GGA TTA TCT CCA CAA   720
Leu Lys Ala Met Gln Val Asp Thr Thr Leu Val Gly Leu Ser Pro Gln
225                 230                 235                 240

GAT GCG AAG AAA TCC CCC TAT ATT GCT TCA ATG GGA GTT TAT GTA TTC   768
Asp Ala Lys Lys Ser Pro Tyr Ile Ala Ser Met Gly Val Tyr Val Phe
                245                 250                 255

AAG ACA GAT GTA TTG TTG AAG CTC TTG AAA TGG AGC TAT CCC ACT TCT   816
Lys Thr Asp Val Leu Leu Lys Leu Leu Lys Trp Ser Tyr Pro Thr Ser
            260                 265                 270

AAT GAT TTT GGC TCT GAA ATT ATA CCA GCA GCT ATT GAC GAT TAC AAT   864
Asn Asp Phe Gly Ser Glu Ile Ile Pro Ala Ala Ile Asp Asp Tyr Asn
        275                 280                 285
```

```
GTC CAA GCA TAC ATT TTC AAA GAC TAT TGG GAA GAC ATT GGA ACA ATT   912
Val Gln Ala Tyr Ile Phe Lys Asp Tyr Trp Glu Asp Ile Gly Thr Ile
    290                 295                 300

AAA TCG TTT TAT AAT GCT AGC TTG GCA CTC ACA CAA GAG TTT CCA GAG   960
Lys Ser Phe Tyr Asn Ala Ser Leu Ala Leu Thr Gln Glu Phe Pro Glu
305                 310                 315                 320

TTC CAA TTT TAC GAT CCA AAA ACA CCT TTT TAC ACA TCT CCT AGG TTC  1008
Phe Gln Phe Tyr Asp Pro Lys Thr Pro Phe Tyr Thr Ser Pro Arg Phe
                325                 330                 335

CTT CCA CCA ACC AAG ATA GAC AAT TGC AAG ATT AAG GAT GCC ATA ATC  1056
Leu Pro Pro Thr Lys Ile Asp Asn Cys Lys Ile Lys Asp Ala Ile Ile
            340                 345                 350

TCT CAT GGA TGT TTC TTG CGA GAT TGT TCT GTG GAA CAC TCC ATA GTG  1104
Ser His Gly Cys Phe Leu Arg Asp Cys Ser Val Glu His Ser Ile Val
        355                 360                 365

GGT GAA AGA TCG CGC TTA GAT TGT GGT GTT GAA CTG AAG GAT ACT TTC  1152
Gly Glu Arg Ser Arg Leu Asp Cys Gly Val Glu Leu Lys Asp Thr Phe
    370                 375                 380

ATG ATG GGA GCA GAC TAC TAC CAA ACA GAA TCT GAG ATT GCC TCC CTG  1200
Met Met Gly Ala Asp Tyr Tyr Gln Thr Glu Ser Glu Ile Ala Ser Leu
385                 390                 395                 400

TTA GCA GAG GGG AAA GTA CCG ATT GGA ATT GGG GAA AAT ACA AAA ATA  1248
Leu Ala Glu Gly Lys Val Pro Ile Gly Ile Gly Glu Asn Thr Lys Ile
                405                 410                 415

AGG AAA TGT ATC ATT GAC AAG AAC GCA AAG ATA GGA AAG AAT GTT TCA  1296
Arg Lys Cys Ile Ile Asp Lys Asn Ala Lys Ile Gly Lys Asn Val Ser
            420                 425                 430

ATC ATA AAT AAA GAC GGT GTT CAA GAG GCA GAC CGA CCA GAG GAA GGA  1344
Ile Ile Asn Lys Asp Gly Val Gln Glu Ala Asp Arg Pro Glu Glu Gly
        435                 440                 445

TTC TAC ATA CGA TCA GGG ATA ATC ATT ATA TTA GAG AAA GCC ACA ATT  1392
Phe Tyr Ile Arg Ser Gly Ile Ile Ile Ile Leu Glu Lys Ala Thr Ile
    450                 455                 460

AGA GAT GGA ACA GTC ATC TGAACTAGGG AAGCACCTCT TGTTGAACTA         1440
Arg Asp Gly Thr Val Ile
465                 470

CTGGAGATCC AAATCTCAAC TTGAAGAAGG TCAAGGGTGA TCCTAGCACG TTCACCAGTT 1500

GACTCCCCGA AGGAAGCTT                                             1519
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 10: FROM 1 TO 470

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Asn Lys Ile Lys Pro Gly Val Ala Tyr Ser Val Ile Thr Thr Glu Asn
 1               5                  10                  15

Asp Thr Gln Thr Val Phe Val Asp Met Pro Arg Leu Glu Arg Arg Arg
            20                  25                  30

Ala Asn Pro Lys Asp Val Ala Ala Val Ile Leu Gly Gly Gly Glu Gly
        35                  40                  45

Thr Lys Leu Phe Pro Leu Thr Ser Arg Thr Ala Thr Pro Ala Val Pro
    50                  55                  60

Val Gly Gly Cys Tyr Arg Leu Ile Asp Ile Pro Met Ser Asn Cys Ile
65                  70                  75                  80

Asn Ser Ala Ile Asn Lys Ile Phe Val Leu Thr Gln Tyr Asn Ser Ala
                85                  90                  95

Pro Leu Asn Arg His Ile Ala Arg Thr Tyr Phe Gly Asn Gly Val Ser
            100                 105                 110

Phe Gly Asp Gly Phe Val Glu Val Leu Ala Ala Thr Gln Thr Pro Gly
        115                 120                 125

Glu Ala Gly Lys Lys Trp Phe Gln Gly Thr Ala Asp Ala Val Arg Lys
    130                 135                 140

Phe Ile Trp Val Phe Glu Asp Ala Lys Asn Lys Asn Ile Glu Asn Ile
145                 150                 155                 160

Val Val Leu Ser Gly Asp His Leu Tyr Arg Met Asp Tyr Met Glu Leu
                165                 170                 175

Val Gln Asn His Ile Asp Arg Asn Ala Asp Ile Thr Leu Ser Cys Ala
            180                 185                 190

Pro Ala Glu Asp Ser Arg Ala Ser Asp Phe Gly Leu Val Lys Ile Asp
        195                 200                 205

Ser Arg Gly Arg Val Val Gln Phe Ala Glu Lys Pro Lys Gly Phe Asp
    210                 215                 220
```

```
Leu Lys Ala Met Gln Val Asp Thr Thr Leu Val Gly Leu Ser Pro Gln
225                 230                 235                 240

Asp Ala Lys Lys Ser Pro Tyr Ile Ala Ser Met Gly Val Tyr Val Phe
                245                 250                 255

Lys Thr Asp Val Leu Leu Lys Leu Leu Lys Trp Ser Tyr Pro Thr Ser
            260                 265                 270

Asn Asp Phe Gly Ser Glu Ile Ile Pro Ala Ala Ile Asp Asp Tyr Asn
        275                 280                 285

Val Gln Ala Tyr Ile Phe Lys Asp Tyr Trp Glu Asp Ile Gly Thr Ile
    290                 295                 300

Lys Ser Phe Tyr Asn Ala Ser Leu Ala Leu Thr Gln Glu Phe Pro Glu
305                 310                 315                 320

Phe Gln Phe Tyr Asp Pro Lys Thr Pro Phe Tyr Thr Ser Pro Arg Phe
                325                 330                 335

Leu Pro Pro Thr Lys Ile Asp Asn Cys Lys Ile Lys Asp Ala Ile Ile
            340                 345                 350

Ser His Gly Cys Phe Leu Arg Asp Cys Ser Val Glu His Ser Ile Val
        355                 360                 365

Gly Glu Arg Ser Arg Leu Asp Cys Gly Val Glu Leu Lys Asp Thr Phe
    370                 375                 380

Met Met Gly Ala Asp Tyr Tyr Gln Thr Glu Ser Glu Ile Ala Ser Leu
385                 390                 395                 400

Leu Ala Glu Gly Lys Val Pro Ile Gly Ile Gly Glu Asn Thr Lys Ile
                405                 410                 415

Arg Lys Cys Ile Ile Asp Lys Asn Ala Lys Ile Gly Lys Asn Val Ser
            420                 425                 430

Ile Ile Asn Lys Asp Gly Val Gln Glu Ala Asp Arg Pro Glu Glu Gly
        435                 440                 445

Phe Tyr Ile Arg Ser Gly Ile Ile Ile Ile Leu Glu Lys Ala Thr Ile
    450                 455                 460

Arg Asp Gly Thr Val Ile
465                 470
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 11: FROM 1 TO 35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTTGATAACA AGATCTGTTA ACCATGGCGG CTTCC 35

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 12: FROM 1 TO 33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCAGTTAAAA CGGAGCTCAT CAGATGATGA TTC 33

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 13: FROM 1 TO 30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTGTGAGAAC ATAAATCTTG GATATGTTAC 30

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 14: FROM 1 TO 28

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAATTCACAG GGCCATGGCT CTAGACCC 28

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 15: FROM 1 TO 40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAGATCAAAC CTGCCATGGC TTACTCTGTG ATCACTACTG 40

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993

(K) RELEVANT RESIDUES IN SEQ ID NO: 16: FROM 1 TO 39

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGGAATTCAA GCTTGGATCC CGGGCCCCCC CCCCCCCCC 39

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 17: FROM 1 TO 24

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGGAATTCAA GCTTGGATCC CGGG 24

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(x) PUBLICATION INFORMATION:
(H) DOCUMENT NUMBER: EP 0536293 A1
(I) FILING DATE: 07-JUN-1991
(J) PUBLICATION DATE: 14-APR-1993
(K) RELEVANT RESIDUES IN SEQ ID NO: 18: FROM 1 TO 32

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCTCTAGACA GTCGATCAGG AGCAGATGTA CG 32

## Claims

1. A method of producing plant seeds having decreased oil content comprising providing increased levels of ADPglucose pyrophosphorylase within said seeds by transforming said plant using the following steps:
(a) inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising
(i) a promoter which functions in plants to cause the production of an RNA sequence in plant seeds,
(ii) a structural DNA sequence that causes the production of an RNA sequence which encodes a

fusion polypeptide comprising an amino-terminal plastid transit peptide and an ADPglucose pyrophosphorylase enzyme,

(iii) a 3' non-translated DNA sequence which functions in plant cells to cause transcriptional termination and the addition of polyadenylated nucleotides to the 3' end of the RNA sequence;

(b) obtaining transformed plant cells; and

(c) regenerating from the transformed plant cells genetically transformed plants which produce seeds having a decreased oil content;

wherein said ADPglucose pyrophosphorylase enzyme is deregulated.

2. The method of claim 1 wherein said enzyme is from *E. coli*.

3. The method of claim 2 wherein said enzyme is *glg*C16.

4. The method of claim 3 wherein said plant is selected from the group consisting of wheat, canola, soybean, corn, cotton, sunflower, almond, cashew, pecan, walnut, and peanut.

European Patent Office

# EUROPEAN SEARCH REPORT

Application Number
EP 94 87 0118

## DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document with indication, where appropriate, of relevant passages | Relevant to claim | CLASSIFICATION OF THE APPLICATION (Int.Cl.6) |
|---|---|---|---|
| A | WO-A-91 19806 (MONSANTO) 26 December 1991<br>* the whole document * | 1-4 | C12N15/82<br>C12N15/54 |
| A | WO-A-93 09237 (SANDOZ) 13 May 1993<br>* the whole document * | 1-4 | |

TECHNICAL FIELDS SEARCHED (Int.Cl.6)

C12N

The present search report has been drawn up for all claims

| Place of search | Date of completion of the search | Examiner |
|---|---|---|
| THE HAGUE | 27 September 1994 | Maddox, A |

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)